R.L. Snyder, *"A Hannawalt Type Phase Identification Procedure for a Minicomputer"*, Adv. In X-ray Analysis, v.24 (1980) pp. 83-90

A HANAWALT TYPE PHASE IDENTIFICATION

PROCEDURE FOR A MINICOMPUTER

Robert L. Snyder

New York State College of Ceramics

Alfred University, Alfred, N.Y. 14802

INTRODUCTION

The use of computers to aid in the identification of phases from their powder diffraction patterns was pioneered in the mid 1960's by Frevel, Nichols and Johnson (1-3). Today's most widely used Johnson algorithm conducts a reverse sequential search by comparing each reference pattern in the JCPDS powder diffraction file (PDF) to the unknown pattern. A figure of merit is computed for each match and the patterns with the best figures of merit are listed at the end of the search. The Nichols approach is a reverse search of a singly inverted reference file. An inverted file is one which stores the reference patterns according to the d value of the lowest angle 100% intensity line (d_1). This type of file is analogous to the Hanawalt search books distributed by the JCPDS for manual searching. When an inverted file is stored in a random format, along with suitable disk directory files, only reference patterns containing d_1 values of interest need be read in the search.

Both the Johnson and Nichols algorithms use the full PDF which today contains about 35000 patterns of highly variable quality (4). The Frevel algorithm is the only one, to date, which attempts to deal with the problem of poor quality reference patterns. This approach also uses a singly inverted reference file; however, the file is restricted to three hundred commonly identified phases in unknowns (5). The use of this drastically restricted file enabled this algorithm to be the first to be converted to run on a laboratory sized minicomputer. This program has been recently generalized by Edmonds. Johnson has also converted his algorithm for use on a minicomputer but due to the exhaustive search approach it also must use a greatly reduced

Advances in X-Ray Analysis, V. 24 (1980) pp. 83-90.

Frevel type of reference file.

Recent developments in computer technology have brought powerful minicomputers with large amounts of mass storage into a price range that most laboratories can afford. This in turn has increased the desirability of a full PDF minicomputer search-match system. For reasons of either mass storage capacity or computational speed none of the existing algorithms are directly convertable to a minicomputer environment. The development of the full PDF search-match system described here was done under the sponsorship of the Siemens Corporation and is currently part of the software supporting their D500 powder diffraction system.

DESIGN CONSTRAINTS

For the implementation of a full file search-match system on a modern minicomputer a number of desirable or essential features must be incorporated. The following design constraints, once established, dictated the form of the search algorithm.

1) The entire data base should be contained in less than five megabytes of disk storage. This allows for implementation on all but the very smallest of current laboratory minicomputer configurations. To meet this constraint a binary compression format was devised for each of the active patterns in the PDF. All d values are converted to integer d-codes by dividing them into 1000. Thus the integers 1 to 2048 will represent d values ranging from 1000 to 0.488. This integer conversion of the d values introduces an average $\Delta 2\theta$ round off error of 0.025°, for Cu radiation d values greater than 1.0, with a maximum round off of 0.05°. Since the average $\Delta 2\theta$ for the cubic patterns in the PDF is 0.1° (4), the maximum round off error of 0.05° does not significantly degrade the reference data. The advantage is that the integers 1 to 2048 can be stored in 11 binary bits.

   Using 11 bits to store the integer d-code leaves five bits which can be used to encode the intensity value and result in the storing of each d-I pair in one 16 bit minicomputer word. Intensities are therefore stored on a scale of 0 to 30. On removing the trailing blanks from the formula the above measures allow for the compression of the 15 megabyte PDF into about 2.5 megabytes.
2) Due to the slow input/output speeds of minicomputers the entire data base should not be searched. This constraint dictates the use of an inverted file with a Hanawalt search strategy. However, a random file structure would violate the file size constraint discussed above. The solution of this dilemma is to create a pseudo-random or indexed sequential file structure. This type of file is inverted by sorting the patterns according to their d_1 values and then the sorted, binary compressed patterns are written sequentially to a disk file. A single disk access will put the program to within 256 words of the

pattern or d_1 range of patterns desired. A double buffered assembly language subroutine performs this function with a negligible amount of time wasted looking at patterns outside the desired d_1 range.

3) Since floating point arithmetic is very slow on most mini-computers it should be avoided. This design contraint is met by first converting the input d values of the unknown pattern into integer d-codes. The entire reverse search is then conducted using only integer arithmetic.
4) Each yearly update to the PDF should be added to the search system with minimum effort. To meet this design constraint each of the sets of the PDF are independently inverted and stored. This produces 28 data files each with an associated d_1 and PDF number directory file.
5) Due to the large volume of patterns in the PDF, and the low probability of most of them being found in common unknowns, a strategy which searches the most likely references first will greatly minimize search time. Following the work of Frevel (5) a file containing approximately 300 of the most common phases in unknowns was created in exactly the same format as described above. This file called the MICRO file is searched first. Any correctly matched phases are quantitatively subtracted from the unknown and only the residual pattern is passed on for further searching. The list of 2500 phases which the JCPDS has designated as frequently encountered have been gathered into a second file called MINI. This file is searched in the second phase of a search. The full 28 set MAXI file is only searched if a residual pattern remains after the MICRO and MINI file searches have been completed.

THE SEARCH ALGORITHM

The hierarchical Hanawalt search proceeds as follows:

1) The d_1 disk block directory of the MICRO file is read into memory.
2) A binary search procedure is used to locate the disk blocks containing the reference patterns whose d_1 values lie within a $\pm$ 0.1° 2θ Cu error window around the d_1 of the unknown.
3) Each reference pattern in the correct d_1 range must pass the following three tests:
   a. Its subfile code (e.g. inorganic, mineral, etc.) must agree with those specified by the user.
   b. All diffraction lines with $I \geq 50$ must at least be present in the unknown.
   c. If the user specified chemical constraints, they must be met.
4) For those patterns which pass the tests a figure of merit (FOM) is computed as described in the next section.
5) The single reference pattern with the highest FOM for d_1 is saved for the MATCH procedure.
6) Steps 2 through 5 are repeated for lines d_2 and d_3.

7) The MATCH routine is where one or more of the three saved patterns may be quantitatively subtracted from the unknown. If a match was found the rescaled residual pattern with a new d_1 value is returned and the search procedure is repeated starting at step 2. If no match is found steps 2 through 7 are repeated with an error window of ± 0.2° 2θ Cu. If this also fails to produce a match the entire procedure is repeated using the next three most intense lines of the unknown (d_4 through d_6) and then again for d_7 through d_9.
8) When no (further) matches are found on searching the 9 highest intensity lines in the (residual) unknown pattern the MICRO file is abandoned. Steps 1 through 7 are then repeated on the MINI file. If any residual pattern remains after the MINI file search, steps 1 through 6 are repeated for each of the 28 files in the MAXI file. After the 28 set MAXI file search is complete 7 is executed, and the process is repeated for lines d_4 through d_9.

The search program accepts data from either the d-I file output by the automated data reduction (ADR) program (6) or from manual entry. It rejects any $K_{\alpha 2}$ lines and only searches patterns in the user selected subfiles. Chemical constraints may be applied.

THE MATCH ALGORITHM

The match routine is entered to evaluate the three patterns found by the search. Any pattern with a figure of merit less than 10 is rejected. If chemistry checking is in effect this minimum acceptable FOM value is lowered to 7.0. The acceptance of an incorrect pattern at this stage will so distort the residual pattern as to make any further correct matches unlikely.

The Figure of Merit contains only three terms:

$$FOM = d_R \times I_R^2 \times d_U$$

d_R = percent of reference lines which match the unknown and which have I>I of the lowest matched line.

I_R = percent of the reference intensity matched.

d_U = percent of the unknown lines matched.

The d_R term does not take into account the closeness of the d agreement. If an unknown line falls within the rather wide error window of the reference line it is called a match. The goodness of fit is only considered when the windows of multiple lines overlap, and then only to match the correct reference and unknown intensities. Since the average Δ2θ Cu for the 2000 indexable cubic PDF patterns is 0.1° (4) and over 1000 are not indexable within a ± 0.5° error window, any term in the FOM which considers goodness of fit beyond the match/no match criterion is not justified.

Explore Litigation Insights

DOCKET ALARM

Docket Alarm provides insights to develop a more informed litigation strategy and the peace of mind of knowing you're on top of things.

Real-Time Litigation Alerts

Keep your litigation team up-to-date with **real-time alerts** and advanced team management tools built for the enterprise, all while greatly reducing PACER spend.

Our comprehensive service means we can handle Federal, State, and Administrative courts across the country.

Advanced Docket Research

With over 230 million records, Docket Alarm's cloud-native docket research platform finds what other services can't. Coverage includes Federal, State, plus PTAB, TTAB, ITC and NLRB decisions, all in one place.

Identify arguments that have been successful in the past with full text, pinpoint searching. Link to case law cited within any court document via Fastcase.

Analytics At Your Fingertips

Learn what happened the last time a particular judge, opposing counsel or company faced cases similar to yours.

Advanced out-of-the-box PTAB and TTAB analytics are always at your fingertips.

API

Docket Alarm offers a powerful API (application programming interface) to developers that want to integrate case filings into their apps.

LAW FIRMS

Build custom dashboards for your attorneys and clients with live data direct from the court.

Automate many repetitive legal tasks like conflict checks, document management, and marketing.

FINANCIAL INSTITUTIONS

Litigation and bankruptcy checks for companies and debtors.

E-DISCOVERY AND LEGAL VENDORS

Sync your system to PACER to automate legal marketing.

WHAT WILL YOU BUILD? | sales@docketalarm.com | 1-866-77-FASTCASE